Infrared Physics & Technology 57 (2013) 56–67

Contents lists available at SciVerse ScienceDirect

Infrared Physics & Technology

journal homepage: www.elsevier.com/locate/infrared

Research on fusion method for infrared and visible images via compressive sensing

Meng Ding [a,*], Li Wei [b], Bangfeng Wang [a]

[a] College of Civil Aviation, Nanjing University of Aeronautics and Astronautics, Nanjing 210016, China
[b] Jin Cheng College, Nanjing University of Aeronautics and Astronautics, Nanjing 210016, China

HIGHLIGHTS

- We use compressive sensing to study fusion method of infrared and visible images.
- This paper firstly proposes the fusion rule of maximum absolute of entry of sparse vector.
- The method using OMP provides better results in the condition of the same parameter setting, dictionary and fusion rule.
- The method ***IRdictionary_maxabsolute_OMP*** takes almost all the largest objective evaluations.

ARTICLE INFO

Article history:
Received 31 October 2012
Available online 21 December 2012

Keywords:
Infrared image
Visible image
Image fusion
Compressive sensing
K-SVD
Sparse vector approximation

ABSTRACT

In order to obtain a more exact, reliable and better description than a single source image, we need to fuse source images taken from different sensors to a synthetic image. This paper employs infrared and visible images and uses the theory of compressive sensing to study image fusion method. The fusion method based on compressive sensing theory contains three parts: overcomplete dictionary, the algorithm of sparse vector approximation and fusion rule. This paper selects three trained overcomplete dictionaries by K-means Singular Value Decomposition (K-SVD) including the dictionary only using patches from the infrared images, the dictionary only using patches from the visible images and the dictionary using the combined patches, two sparse vector approximations containing orthogonal matching pursuit and polytope faces pursuit algorithms, and two fusion rules covering maximum ℓ^1-norm and maximum absolute of entry of sparse vector which is firstly proposed in this paper to study twelve fusion approaches. The experimental results show that the method using orthogonal matching pursuit can provide better fusion results in the condition of the same parameter setting and the same dictionary and fusion rule, and the method using the dictionary only using patches from the infrared images, the fusion rule of maximum absolute of entry of sparse vector and orthogonal matching pursuit takes almost all the largest objective evaluations and the best fusion quality.

© 2012 Elsevier B.V. All rights reserved.

1. Introduction

In recent years, many images for one object or scene can be acquired by multiple sensors with the technology development of image sensors. In accordance with the natural properties of the sensors and the approach the images are obtained, all of these source images taken from the sensor directly contain unique information which is usually complementary. Taking infrared and visible images studied in this paper for example, infrared images have lower contrast and definition compared with visible images, but visible images cannot capture targets effectively in low visibility conditions [1,2]. The fusion of infrared and visible images can obtain a more exact, reliable and better description than a single source image. Therefore, we can think that image fusion is the foundation of the analysis of multisensor images [3]. In order to extend or enhance information about the scene, it is necessary to develop the technology of image fusion by combining the images captured by different sensors. Image fusion is the process of detecting salient features in the source images and fusing these details to a synthetic image. For the past few years, image fusion has been applied widely in diverse fields, such as target detection, intelligent surveillance, and nondestructive inspection [4–6].

All of image fusion methods developed or applied in the past two decades can be divided into three levels: pixel-level, feature-level, and decision-level in accordance with the stage where the information acquired by different image sensors is fused to a synthetic [7,8]. Pixel-level fusion method which combines the independent source images into a single image, reserves most of the information and is studied widely. Feature-level methods which typically use features of source images (such as edges or regions) to fuse them are usually robust to noise and misregistration. Since

* Corresponding author.
E-mail address: nuaa_dm@hotmail.com (M. Ding).

1350-4495/$ - see front matter © 2012 Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.infrared.2012.12.014

decision-level methods combine image descriptions directly, the application field of them is limited greatly. The method proposed in this paper belongs to pixel-level fusion. The pixel-level approaches mainly contain two categories [9,10]: spatial domain-based methods and transformed domain-based methods. The simplest fusion method in spatial domain just takes the pixel-by-pixel average of the source images. Spatial domain-based methods often lead to undesirable side effects, such as lower contrast, blockness in the fused image [11]. In the past decades, one of the most successful image fusion methods is by using multiscale transform [12]. As an outstanding transformed domain-based method, the image fusion method based on multiscale transform has three basic steps [13]: (1) decompose source images into multiscale representations with different resolutions and orientations; (2) decompose coefficients denoting multiscale representations and linked with the salient features of the original images are integrated according to fusion rules; (3) reconstruct the fused image by the inverse transform of the integrated multiscale coefficients.

The earliest multiscale transform for image fusion is pyramid decomposition, for example, Laplacian pyramid [14], morphological pyramid [15], gradient pyramid [16]. Pyramid method firstly decomposes each source image into a series of images with different sizes (calls pyramid) in different resolutions, and then extracts the value in the pyramid with the highest saliency at each position in the decomposed images, finally reconstructs the fused image using the inverse transform of the composite images. Another multiscale transform for image fusion is wavelet transform-based methods which use a similar scheme to the pyramid decomposition, such as discrete wavelet transform [17,18], stationary wavelet transform [19], and dual-tree complex wavelet transform [20,21]. The principal shortcoming of these methods is that most of the multiscale transforms are not shift invariant, which is brought by the underlying down-sampling process. Recently, multiscale geometry analysis has been developed and used for image fusion to improve fusion results. Typical methods include ridgelet transform [22], curvelet transform [23]. However, although different wavelets can represent different image details, wavelets and related multiscale transform cannot extract all of the underlying information of the source images effectively. The reason is that the dictionary constructed by different basis functions is limited. Decomposed coefficients of the fused image obtained by a limited dictionary in the transformed domain may cause all the pixel values to change in the spatial domain. Consequently, in some cases multiscale transform-based fusion methods may produce undesirable artifacts.

Obviously, in order to make the fused image more accurate, it is necessary to explore a novel method to extract the underlying information of the source images more efficiently and completely. This paper proposes an image fusion method which is based on the recently developed theory of compressive sensing (CS). CS theory has been successfully applied in different fields of image process or computer vision [24–27], such as image denoising, image compression, feature extraction, and target classification. The core of CS is the sparse representation which describes natural signals including images by a sparse linear combination of columns of an overcomplete dictionary. Different from a limited dictionary within multiscale transformations, CS uses an overcomplete dictionary in which every column is also called a signal atom. Overcompleteness is the most prominent characteristic of the dictionary used in CS theory. Overcompleteness denotes that the number of signal atoms in the overcomplete dictionary is more than signal dimensions greatly and guarantees more meaningful and complete representation of source signals than the traditional multiscale transformations [28]. CS theory reveals the coefficients corresponding to the natural sign are sparse. Thus, Li and Yang employ the sparse coefficient vectors to give a framework of CS-based image fusion [29]. According to this framework, our method for image fusion proposed in this paper contains following steps: Firstly, the overcomplete dictionary is created and trained by K-means Singular Value Decomposition (K-SVD) [30]. Secondly, the source images are divided into patches by sliding window which is adopted to achieve better performance in capturing local salient features of source images and improve the image fusion quality. Thirdly, the patches are decomposed by the overcomplete dictionary into their corresponding sparse coefficients. Fourthly, tow fusion rules are employed to combine the coefficients of the source images. Finally, the fused image is reconstructed using the combined coefficients.

The rest of the paper is organized as follows: Section 2 presents the basic theory of CS and K-SVD for the overcomplete dictionary creation. In Section 3, the fusion scheme based on CS and the fusion rules based on sparse coefficients are discussed. Experiment and conclusion are demonstrated in Sections 4 and 5 respectively.

2. Basic theory of compressive sensing

In order to finish image fusion using CS-based method, a brief description of CS theory is necessary. CS theory has four essential factors. First, as mentioned above, overcompleteness is one of major characteristics of CS. Additional important conception is sparsity. Third, how to compute sparse coefficients denoting linear combinations of overcomplete dictionary is a problem. Lastly, creating an overcomplete dictionary is the key step of using CS theory in practices.

2.1. Overcomplete representation

Suppose a given signal vector $\mathbf{y} \in \mathcal{R}^{\boldsymbol{n}}$, and a collection of vectors $\boldsymbol{\varphi}_i \in \mathcal{R}^{\boldsymbol{n}}$, $i = 1, \ldots, m$, where $m > n$. Such collections are usually dictionaries and each vector $\boldsymbol{\varphi_i}$ is an atom. Given signal $y = \sum a_i \boldsymbol{\varphi_i}$ ($i = 1, \ldots, m$) can be represented as a linear combination of atoms in the dictionary. Different from traditional multiscale transform basis representation, such linear combination of dictionary offers a wider range of generating atoms [31]. Thus, this dictionary is overcomplete and called overcomplete dictionary [32]. Overcomplete representation based on such dictionary can allow more flexibility in signal representation and more effectiveness in signal process [33].

Considering the atoms as the columns of overcomplete dictionary $\boldsymbol{\Phi}$, overcomplete dictionary $\boldsymbol{\Phi} = [\boldsymbol{\varphi_1}, \boldsymbol{\varphi_2}, \ldots, \boldsymbol{\varphi_m}]$, so that the matrix $\boldsymbol{\Phi} \in \mathcal{R}^{n \times m}$. Linear algebra tells us that a representation of the given signal $\mathbf{y}$ can be described as a coefficient vector $\mathbf{a} = [a_1, a_2, \ldots, a_m]^{\mathbf{T}}$ satisfying $\mathbf{y} = \boldsymbol{\Phi}\mathbf{a}$. Since $m > n$, the problem of overcomplete representation is undetermined. That means there is no unique solution of the coefficients vector $\mathbf{a}$. In order to obtain unique solution, considering the impact of sparsity constraint on this situation, CS theory can in certain circumstances generate a sparse coefficient vector as the linear combination of overcomplete dictionary. This coefficient vector is called sparse representation (shown in Fig. 1).

2.2. Sparse representation and sparse vector approximation

Generally speaking, the purpose of CS theory is to solve the problem of finding the sparsest representation possible in an overcomplete dictionary. As a measure of sparsity of a vector $\mathbf{a}$, ℓ^0-*norm* $\|\mathbf{a}\|_0$ denotes the number of non-zero entries in $\mathbf{a}$. The sparsest representation is the solution to the optimization problem [34]:

$$\min_{\mathrm{a}} \|\mathrm{a}\|_0 \qquad \text{s.t.} \qquad \mathrm{y} = \boldsymbol{\Phi}\mathrm{a} \tag{1}$$

In most practical situations, the above formula can be modified to include a noise allowance [33]:

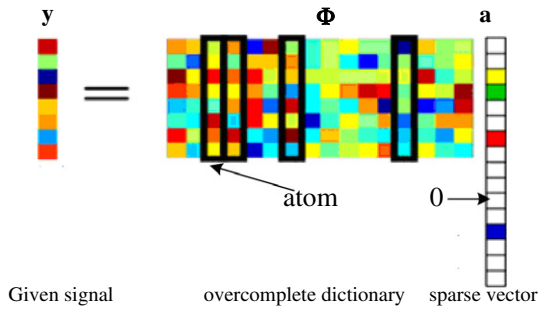


Fig. 1. Overcomplete dictionary and sparse representation.

$$\min_{a}\|a\|_0 \qquad \text{s.t.} \qquad \|y-\Phi a\|_2 \leqslant \delta \tag{2}$$

The above optimization is an NP-hard problem. In order to looking for the sparsest representation of the vector **y**, we have to enumerate all possible combinations of columns of **Φ** Generally, such an algorithm has to cost at $\boldsymbol{O}(2^m)$ flops to carry out [33]. Thus, the approximations or relaxations of this optimization problem have to be researched to replace computation of analytical solutions.

In the past decade, several algorithms have been proposed to solve this problem. The most typical methods contain a series of greedy algorithms and l^1-*norm* minimization-based algorithms. As the simplest greedy algorithm, Matching Pursuit (MP) builds up k-atom ($k \leqslant m$) approximate representations a step in an iteration, adding to an existing ($k-1$)-atom approximation a new term chosen in a greedy fashion to minimize the resulting l^2-*norm* error [31]. When stopped after N iterations, one gets a sparse approximate representation. Based on MP, Orthogonal MP (OMP) is further proposed to reduce iterations and reconstruction error, improve the robustness [35–37]. The basic idea of l^1-*norm* minimization-based algorithm is to replace the l^0-*norm* with a l^1-*norm*. Recent development in the emerging theory of applied mathematics reveals that the solution of the l^0-*norm* minimization problem is equal to the solution to the following l^1-*norm* minimization problem in under certain conditions [34]:

$$\min_{a}\|a\|_1 \qquad \text{s.t.} \qquad y=\Phi a \tag{3}$$

Considering the noise allowances, the above formula is modified to the following formula:

$$\min_{a}\|a\|_0 \qquad \text{s.t.} \qquad \|y-\Phi a\|_2 \leqslant \delta \tag{4}$$

This l^1-*norm* minimization problem can be solved by standard linear programming methods In CS theory, this approach replacing the l^0-*norm* with a l^1-*norm* is called Basis pursuit (BP) [38,39] Because the solution is known to be very sparse, even more efficient methods based on BP algorithm can be available, such as Polytope Faces Pursuit (PFP) algorithm [40].

2.3. Creation of overcomplete dictionary

For image fusion, given signal **y** can be obtained directly from source images, but the overcomplete dictionary **Φ** cannot. Therefore, the technique of designing dictionary to better fit sparse representation has been considered and discussed. At present, these techniques mainly divided into two categories: the first category is that the dictionary is selected from overcomplete transform bases or be created as a hybrid of several multiscale transform bases. The typical examples of this category include overcomplete Discrete Cosine Transform (DCT) dictionary consisting of DCT bases and the hybrid dictionary consisting of DCT bases, wavelet bases, ridgelet bases and so on [29]. The other category is a trained dictionary obtained from learning train samples. K-SVD present in paper [30] belongs to the second category. Recent studies reveal that the dictionary trained by K-SVD outperforms the dictionary of the first category in the field of image process, such as image reconstruction and image denoising [24].

The idea of K-SVD comes from K-means for Vector Quantization (VQ). Typically, the K-means algorithm is used to train the dictionary of VQ codewords. Suppose codebook matrix $\mathbf{C}=[\mathbf{c}_1, \mathbf{c}_2, \ldots, \mathbf{c}_K]$, $\mathbf{c}_i$ is the codeword of codebook(dictionary), when **C** is given, each signal $\mathbf{y}_i$ is represented by a vector as its closest codeword under ℓ^2-*norm* distance. That means $\mathbf{y}_i=\mathbf{C}\mathbf{a}_i$, where $\mathbf{a}_i=\mathbf{e}_j$ is a vector from the trivial basis with all zero entries except a one in the jth position [41]. The K-means can be considered as an extreme case of sparse representation in the sense that only one atom (codeword) is allowed to participate in the linear combination of $\mathbf{y}_i$. The representation Mean Square Error (MSE) of $\mathbf{y}_i$ is defined as $\|\mathbf{y}_i-\mathbf{C}\mathbf{a}_i\|_2^2$, and for a wide family of signals $\mathbf{Y}=\{\mathbf{y}_i\}_{i=1}^N (N>>K)$, the overall MSE is $\|\mathbf{Y}-\mathbf{C}\mathbf{A}\|_F^2$ where $\mathbf{A}=\{\mathbf{a}_i\}_{i=1}^N$. Therefore, the objective function of K-means is as follows:

$$\min_{\mathbf{C},\mathbf{A}}\|\mathbf{Y}-\mathbf{C}\mathbf{A}\|_F^2 \qquad \text{s.t} \qquad \forall i, \quad \mathbf{a}_i=\mathbf{e}_k \text{ for some } k \tag{5}$$

The K-means algorithm is an iterative method. Each iteration contains two stages, one is computing **A** and the other is updating the codebook. In K-means algorithm, each given signal must be represented by only one atom (codeword). When a few atoms of the dictionary can be allowed to represent a given signal, K-means algorithm transform to K-SVD algorithm. In other words, the coefficient vector denoting a linear combination of a given signal is allowed more than one non-zero entry in K-SVD, and K-SVD can be viewed as a generalization of K-means. For this case, the optimization problem corresponding to Eq. (5) is that of searching the best possible dictionary **D** for the sparse representation of the given signals set **Y**

$$\min_{\mathbf{D},\mathbf{A}}\|\mathbf{Y}-\mathbf{D}\mathbf{A}\|_F^2 \qquad \text{s.t.} \qquad \forall i, \quad \|\mathbf{a}_i\|_0 \leqslant T_0 \tag{6}$$

where the dictionary **D** is similar the codebook **C** of K-means algorithm, T_0 is a threshold denoting the allowed maximization number of non-zero entries of the coefficient vector.

K-SVD algorithm is used to obtains optimal solution of Eq. (6) by iterative approach. One iterative process can be divided into two stages, sparse representation computation stage and dictionary update stage. In the first stage, the algorithm fixes **D** and uses the algorithm of sparse vector approximation to obtain the coefficient matrix. The purpose of the second stage is to search for a better dictionary. The process updates only one column of **D** at a time. That means fixing all columns of **D** except one $\mathbf{d}_k$ and finding a new atom $\tilde{\mathbf{d}}_k$ and new values for its coefficients that best reduce the MSE. Singular Value Decomposition (SVD) is employed to decompose the overall representation error matrix $\mathbf{E}_k$ which stands for the error for all N columns vector of **Y** when the kth atom is removed and $\mathbf{E}_k=\mathbf{Y}-\sum \mathbf{d}_j\mathbf{a}_T^j$, $j=1, 2, \ldots, k-1, k+1, \ldots, N$, $\mathbf{a}_T^j$ denotes the jth row in coefficient matrix **A** that correspond to one column $\mathbf{d}_j$ in the dictionary. The stopping condition of iteration can be set to the given iteration number.

In this paper, we select three different trained dictionaries to study image fusion: (1) the dictionary by only training patches from the infrared image; (2) the dictionary by only training patches from the visible images; (3) the dictionary by training on a corpus of patches taken from infrared and visible images.

3. Image fusion algorithm based on CS theory

Generally speaking, source images used for fusion must be registered. Recently, many effective approaches for image registration have been proposed. The source images used in this paper from

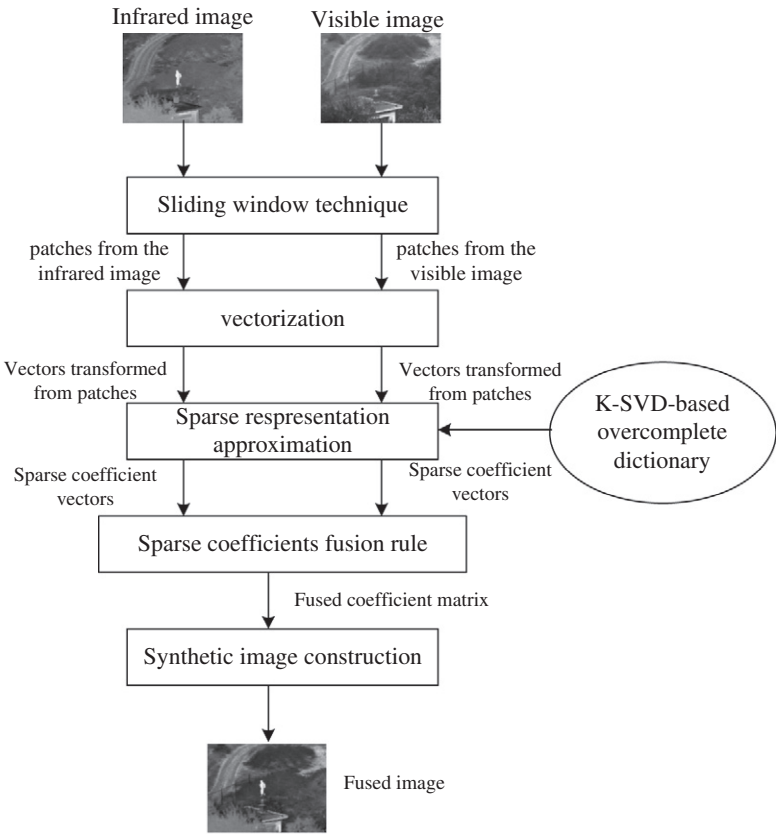


Fig. 2. Framework of image fusion based on CS theory.

www.imagefusion.org have been geometrically registered. The framework of image fusion based on CS theory is summarized in Fig. 2. The method mainly contains three stages: the first stage is to obtain given signal set $\mathbf{Y}$ from source images using sliding window technique and image patch vectorization. In the second stage, sparse coefficient matrixes of different source images $\mathbf{I}^{in}$ and $\mathbf{I}^{vi}$ are computed by approximations including OMP and l^1-*norm* minimization-based algorithm. Thirdly, fused sparse coefficient matrix $\mathbf{A}^{F}$ is combined using coefficients fusion rule. Lastly, construct fused image $\mathbf{I}^{F}$ in accordance with fused sparse coefficient matrix $\mathbf{A}^{F}$.

3.1. Obtain given signal set $\boldsymbol{Y}$

It is obvious that a signal set $\mathbf{Y} = \{\mathbf{y}_i\}_{i=1}^{N}$ comes from source images and a source image corresponds to one given signal set $\mathbf{Y}$. Because image fusion generally depends on the local information of source images, sparse coefficient matrix representing the entire image cannot directly be used with image fusion. In order to solve this problem, Li proposed a method that divides the source images into small patches and makes the sparse representation shift invariant by a sliding window technique [9,29]. Sliding window technique and the method to obtain the given signal set $\mathbf{Y}$ are shown in Fig. 3. Firstly, divide source image $\mathbf{I}$ of size $M \times N$ from left-top to right-bottom into every possible image patches of size $n \times n$, where n is the length of atom in the dictionary and $n < M$, $n < N$, and then each image patch is vectorized by row-to-column order to the column $\mathbf{y}_i$ of $\mathbf{Y}$. when the sizes of the source image and the image patch are $M \times N$ and $n \times n$ respectively, the size of given signal set $\mathbf{Y}$ is $n^2 \times (M - n + 1)(N - n + 1)$. The size of image patch (sliding window) plays an important role in CS-based image

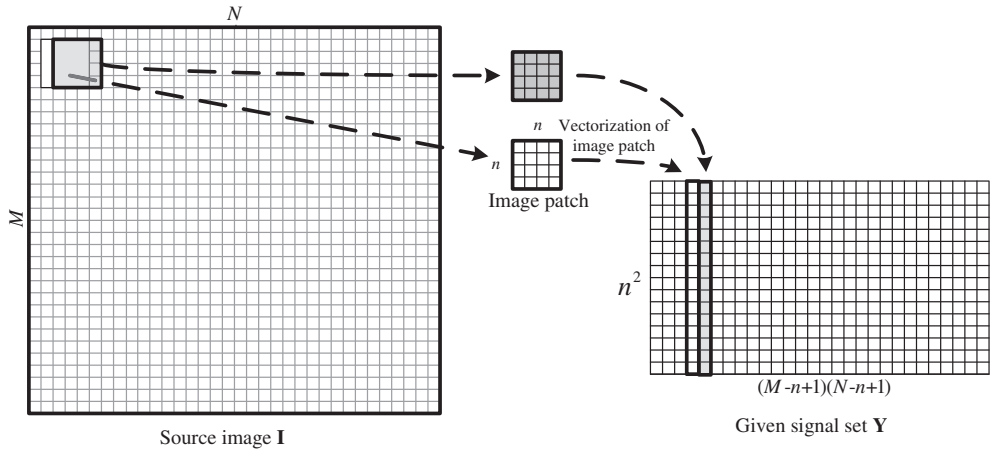


Fig. 3. Process of obtaining given signal set $\mathbf{Y}$.

fusion [9]. The larger size patches led to the length of the atom of the corresponding dictionary increases and the calculation speed of sparse vectors reduces. As the size of the patch decreases, the process of sparse vectors computation becomes faster, but the information content contained in the patches would be lower and it may miss some of the important features of the source images. In this paper, the size of the image patch n is equal to 8 which has been proved to be appropriate setting for the image denoising and fusion [9,24].

3.2. Fusion scheme

We assume that the source images $\mathbf{I}^{in}$ and $\mathbf{I}^{vi}$ taken from infrared and visual sensor respectively. As mentioned earlier, given signal sets $\mathbf{Y}^{in}$ corresponding to $\mathbf{I}^{in}$ and $\mathbf{Y}^{vi}$ corresponding to $\mathbf{I}^{vi}$ are obtained by using the slide window technique. In matrix $\mathbf{Y}^{in}$ or $\mathbf{Y}^{vi}$, each column $\mathbf{y}_i^{in}$ or $\mathbf{y}_i^{vi}$ corresponds to one patch of the source image $\mathbf{I}^{in}$ or $\mathbf{I}^{vi}$, $\mathbf{a}_i^{in}$ and $\mathbf{a}_i^{vi}$ are the sparse representation of each column $\mathbf{y}_i^{in}$ and $\mathbf{y}_i^{vi}$ respectively. The process of image fusion based on sparse representation matrix is as follows:

- *Step. 1*. According to the sparse coefficient fusion rule, corresponding columns of sparse representation matrixes $\mathbf{A}^{in}$ and $\mathbf{A}^{vi}$ of the source images are fused to generate fused sparse representation matrix $\mathbf{A}^F$;
- *Step. 2*. The vector representation of the fused image $\mathbf{Y}^F$ can be obtained by $\mathbf{Y}^F = \mathbf{\Phi}\mathbf{A}^F$;
- *Step 3*. As the inverse process of obtaining given signal set, reshape each column $\mathbf{y}_i^F$ of $\mathbf{Y}^F$ into a block with the size $n \times n$ and then add the block to the null matrix $\mathbf{S}$ with the size of $M \times N$;
- *Step 4*. According to Fig. 3, in each pixel position of $\mathbf{S}$, the pixel value is the sum of several block values, thus, the pixel value of $\mathbf{S}$ is divided by the adding times at its position to obtain the corresponding pixel value of fused image $\mathbf{I}^F$. The adding times in each pixel position of $\mathbf{S}$ are indicated as follows:

$$\text{adding times} = \begin{bmatrix} 1 & 2 & 3 & \cdots & n & n & n & \cdots & 3 & 2 & & 1 \\ 2 & 4 & 6 & \cdots & 2n & 2n & 2n & \cdots & 6 & 4 & \iddots & \vdots \\ n & 2n & 3n & \cdots & n^2 & n^2 & n^2 & \cdots & 3n & 2n & & n \\ n & 2n & 3n & \cdots & n^2 & n^2 & n^2 & \cdots & 3n & 2n & & n \\ n & 2n & 3n & \cdots & n^2 & n^2 & n^2 & \cdots & 3n & 2n & \iddots & \vdots \\ 2 & 4 & 6 & \cdots & 2n & 2n & 2n & \cdots & 6 & 4 & & 2 \\ 1 & 2 & 3 & \cdots & n & n & n & \cdots & 3 & 2 & & 1 \end{bmatrix}$$

The coefficient fusion rule is usually determined by the activity level of the atom which can be described by the absolute value of the corresponding coefficient in the sparse representation vectors [10,13]. Several algorithms employ averaging $\bar{a}_i$ or absolute maximum $\|\mathbf{a}_i\|_1$ to express the activity level. The averaging fusion rule denotes that the corresponding coefficients are averaged to obtain fused sparse vector. This rule can retain the contrast information of the source images, but the high-frequency features denoting details would get smoother. Generally, fusion algorithm always hopes to integrate the information of the source images into fused image as more as possible. This paper studies two rules to combine the coefficient vectors. The first is called *maximum-l^1-norm* fusion rule, which obtains the fused coefficient vector by selecting the coefficient vector with maximum ℓ^1-norm [29]. Because the maximum coefficient absolution of the sparse vector denotes the most important information of source images, this paper proposes the second fusion rule calls *absolute of entry maximum* (*maxabsolute*) fusion rule, which obtains the fused coefficient vector by choosing the coefficient vector with maximum entry absolute of the sparse vector and is calculated by

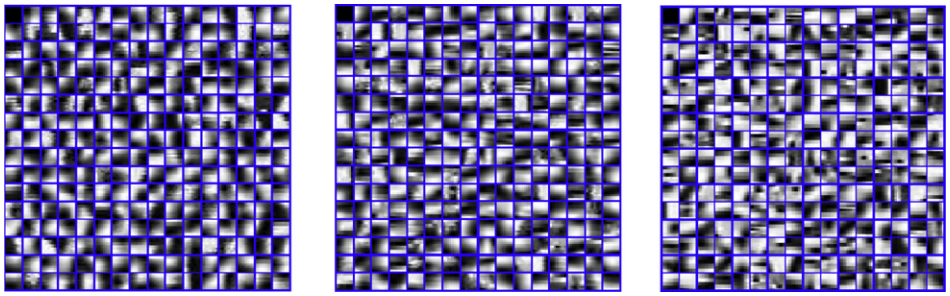
$$\mathbf{a}_i^F = \begin{cases} \mathbf{a}_i^{in} & \text{if } \max\left(|\mathbf{a}_i^{in}|\right) \geqslant \max\left(\mathbf{a}_i^{vi}|\right) \\ \mathbf{a}_i^{vi} & \text{if } \max\left(|\mathbf{a}_i^{in}|\right) \geqslant \max\left(\mathbf{a}_i^{vi}|\right) \end{cases} \tag{7}$$

where $\mathbf{a}_i^{in}$ and $\mathbf{a}_i^{vi}$ are the sparse representation, $|\cdot|$ is the absolute of each entry of vector, $\max(\cdot)$ is equal to the maximum entry of the vector.

4. Experimental verification and discussion

4.1. Experiment setup

This section mainly demonstrates the fused images and compares objective evaluation measures of fusion results in different overcomplete dictionaries, different fusion rules and different sparse vector approximations.



(a) trained dictionary only using patches from the infrared image (b) trained dictionary only using patches from the visible images (c) hybrid dictionary with a corpus of patches from infrared and visible

Fig. 4. Three different overcomplete dictionaries.

The overcomplete dictionaries used in this paper are trained by the K-SVD algorithm. Training samples consist of 20000 patches of size 8×8 which randomly take from different images including infrared and visible images. According to the proportion of different types of training patches, three kinds of overcomplete dictionaries are used to test and evaluate the performance of the proposed methods. Fig. 4a–c respectively show the trained dictionaries only using patches from the infrared images (*IRdictionary*), only using patches from the visible images (*VIdictionary*) and using the combined patches consisting of half from infrared and half from visible images (*HYdictionary*). In Fig. 4, K of K-SVD algorithm is set to 256 and iteration number of the dictionary training is equal to100.

Because source images are acquired by various sensors, and for different types of source images, the performances of fused images using a fusion algorithm are not identical, thus, it is a difficult task to evaluate fused image quality and there are several different criterions according to different purposes. The fusion evaluation criterions contain objective and subjective measures. The subjective measure is to observe fused images with the eye. An objective fusion measure should extract all the perceptually important information that exists in the input images and measure the ability of the fusion process to transfer as accurately as possible this information into the fused image. For the fusion of infrared and visible images discussed in this paper, we select three objective evaluation measures which have been proved to be validated in large degree and considered to quantitatively evaluate the fusion performances [42], including **MI** (mutual information), $\boldsymbol{Q_W}$, and $\boldsymbol{Q^{AB/F}}$. The larger the value of **MI** indicates better fusion performance. For the outstanding fusion result, $\boldsymbol{Q^{AB/F}}$ should be as close to 1 as possible. The larger $\boldsymbol{Q_W}$ denotes the better fused results.

As mentioned earlier, according to the selection of three different overcomplete dictionaries, two fusion rules of coefficient matrix and two sparse vector approximations, the methods discussed and compared in this paper contains:

(1) ***IRdictionary_maxl^{1}norm_OMP***: This method employs trained dictionary only using patches from the infrared image, l^1-norm maximum fusion rule, and OMP sparse vector approximation.

(2) ***IRdictionary_maxl^{1}norm_PFP***: This method employs trained dictionary only using patches from the infrared image, l^1-norm maximum fusion rule, and PFP sparse vector approximation.

(3) ***VIdictionary_maxl^{1}norm_OMP***: This method employs trained dictionary only using patches from the visible image, l^1-norm maximum fusion rule, and OMP sparse vector approximation.

(4) ***VIdictionary_maxl^{1}norm_PFP***: This method employs trained dictionary only using patches from the visible image, l^1-norm maximum fusion rule, and PFP sparse vector approximation.

(5) ***HYdictionary_maxl^{1}norm_OMP***: This method employs trained dictionary using patches from the hybrid image, l^1-norm maximum fusion rule, and OMP sparse vector approximation.

(6) ***HYdictionary_maxl^{1}norm_PFP***: This method employs trained dictionary using patches from the hybrid image, l^1-norm maximum fusion rule, and BP sparse vector approximation.

(7) ***IRdictionary_maxabsolute_OMP***: This method employs trained dictionary only using patches from the infrared image, absolute of entry maximum fusion rule, and OMP sparse vector approximation.

(8) ***IRdictionary_maxabsolute_PFP***: This method employs trained dictionary only using patches from the infrared image, absolute of entry maximum fusion rule, and PFP sparse vector approximation.

(9) ***VIdictionary_maxabsolute_OMP***: This method employs trained dictionary only using patches from the visible image, absolute of entry maximum fusion rule, and OMP sparse vector approximation.

(10) ***VIdictionary_maxabsolute_PFP***: This method employs trained dictionary only using patches from the visible image, absolute of entry maximum fusion rule, and PFP sparse vector approximation.

(11) ***HYdictionary_maxabsolute_OMP***: This method employs trained dictionary using patches from the hybrid image, absolute of entry maximum fusion rule, and OMP sparse vector approximation.

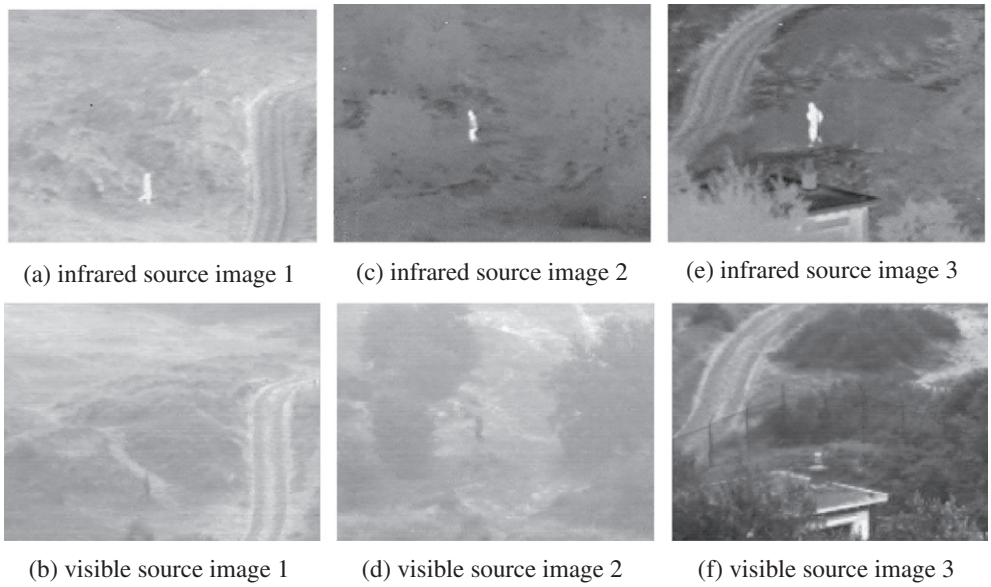


Fig. 5. Source images.

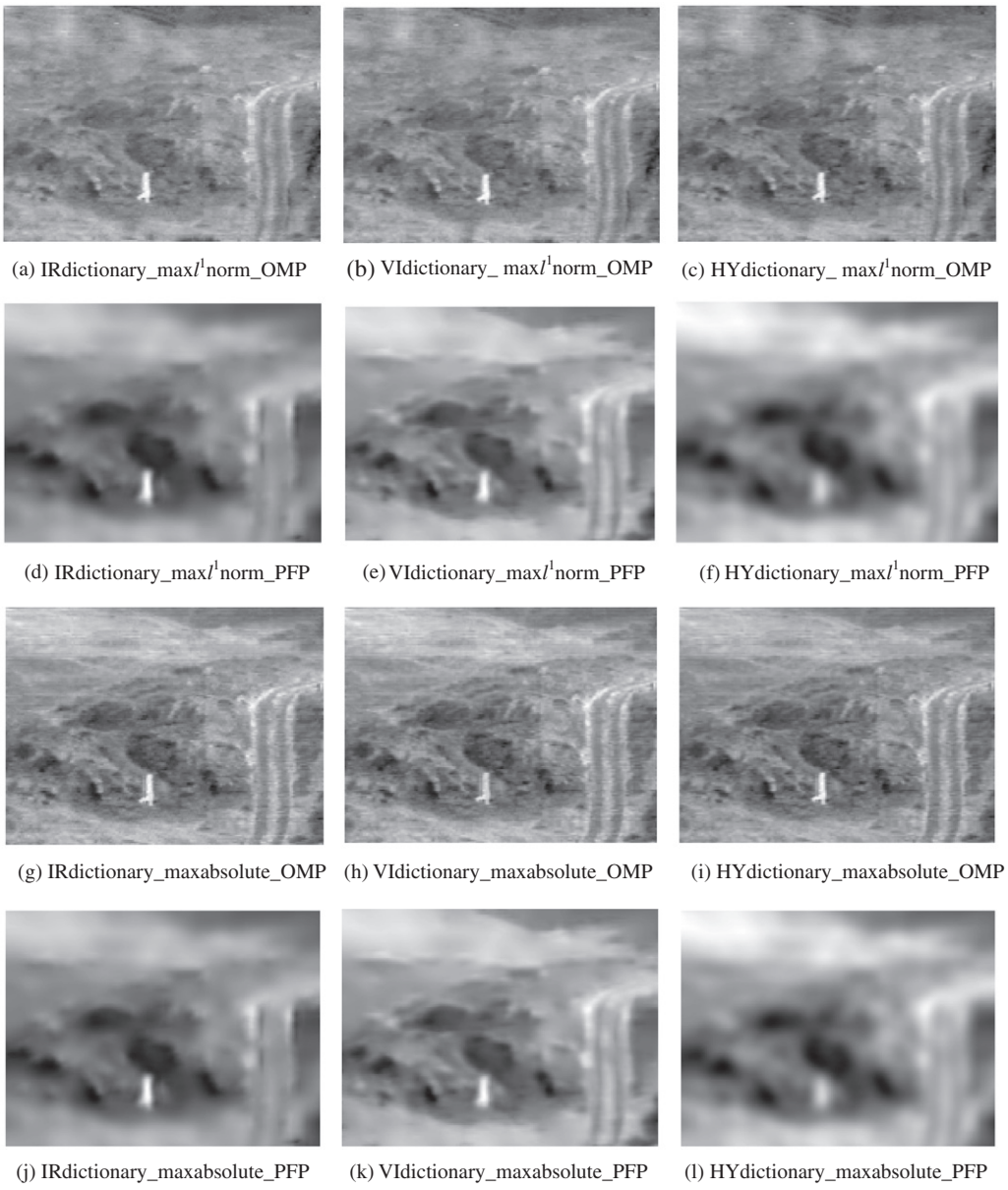


Fig. 6. Fused images of source images 1 using different approaches.

(12) ***HYdictionary_maxabsolute_PFP***: This method employs trained dictionary using patches from the hybrid image, absolute of entry maximum fusion rule, and PFP sparse vector approximation.

In the experiment, three pairs of infrared and visible source images, shown in Fig. 6, are used to test 12 methods. In infrared source images, such as Fig. 5a, c, and e, the person can be observed clearly, while in visible source images, such as Fig. 5b, d, and f, the background including trees, roads and barriers is clear. Fused algorithm needs to show the person and background clearly in a fused image.

4.2. Experiment results and performance evaluation

The fused images of Fig. 5a and b with 12 different fusion approaches are shown in Fig. 6a∼l. Fig. 7a∼l and Fig. 8a∼l demonstrate the fused results of source image pair 2 and 3 of Fig. 5 respectively. In these experiments, we fix the global error δ in Eqs. (2) and (4) used in OMP and PFP algorithm to 1, the size of all overcomplete dictionaries shown in Fig. 4 is 64×256. Figs. 6–8 visually show that the visual quality of the first and third rows (a, b, c, g, h, and i) using OMP sparse vector approximation is better than the second and fourth rows (d, e, f, j, k, and l) using PFP approximation in the same parameter setting. For example, the fused images of the infrared and visible images in the second and fourth rows of Figs. 6–8 are blurred and the details and features of the source images cannot be reserved availably. On the contrary, the fused images of the first and third rows of Figs. 6–8 faithfully reserve important detailed information of source images. In these images the person are clearly enhanced and some other useful background information is almost preserved. From Figs. 6–8, we can see that the methods using OMP approximation can provide better visual fusion results in the condition of same parameter setting.

The objective evaluations containing **MI**, $\boldsymbol{Q_W}$, and $\boldsymbol{Q^{AB/F}}$ on the fused images of Figs. 6–8 are listed in Table 1–3 respectively. According to these tables, we also find that the objective

evaluations of all methods using OMP approximation (a, b, c, g, h, and i columns) are obviously better than all algorithms using PFP approximation (d, e, f, j, k, and l columns). The method *VIdictionary_maxl1norm_PFP* corresponding to e column of Table 1–3 is the best in all methods using PFP approximation. Thus, in the next experiment, we employ this method to discuss the relationship between the global error δ in Eq. (4) and the fusion quality. The method *IRdictionary_maxabsolute_OMP* corresponding to g column of Table 1–3 takes almost all the largest objective evaluations which indicated in bold. Therefore, in the following experiment, we select this method to study the relationship between the objective evaluations and the length of trained overcomplete dictionary (the value of K) and iterations of dictionary training. From Table 1–3, we also can find that the different fusion rules have no effect on the objective evaluations (d–f columns to j–l columns) in the different dictionaries using PFP approximation.

The global error δ in Eq. (4) has an important effect on the speed of PFP algorithm computation. As the value of δ decreases, the process of the PFP becomes slower quickly. However, careful inspection of Fig. 9 reveals that visual quality of the fused image has been improved obviously as the value of δ decreases. The objective evaluations of Fig. 9 shown in Table 4 can also verify this conclusion. Compared with the best objective evaluations of Table 3, the best values of Table 4 indicated in bold are not perfect. Therefore, in general the performance of OMP is better than PFP in image fusion.

The number of atoms of overcomplete dictionary K and the iteration for dictionary training are two important parameters. The

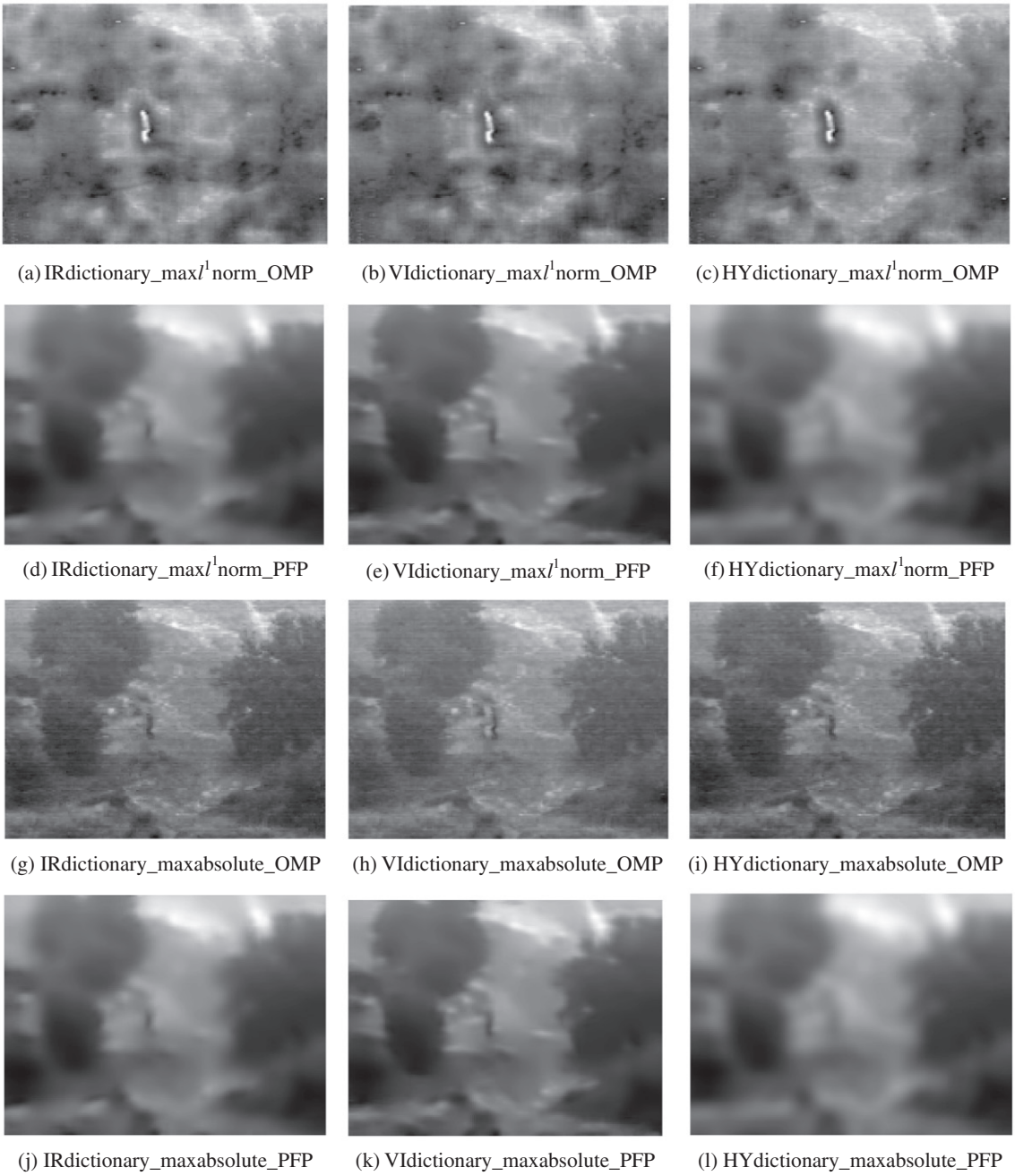


Fig. 7. Fused images of source images 2 using different approaches.

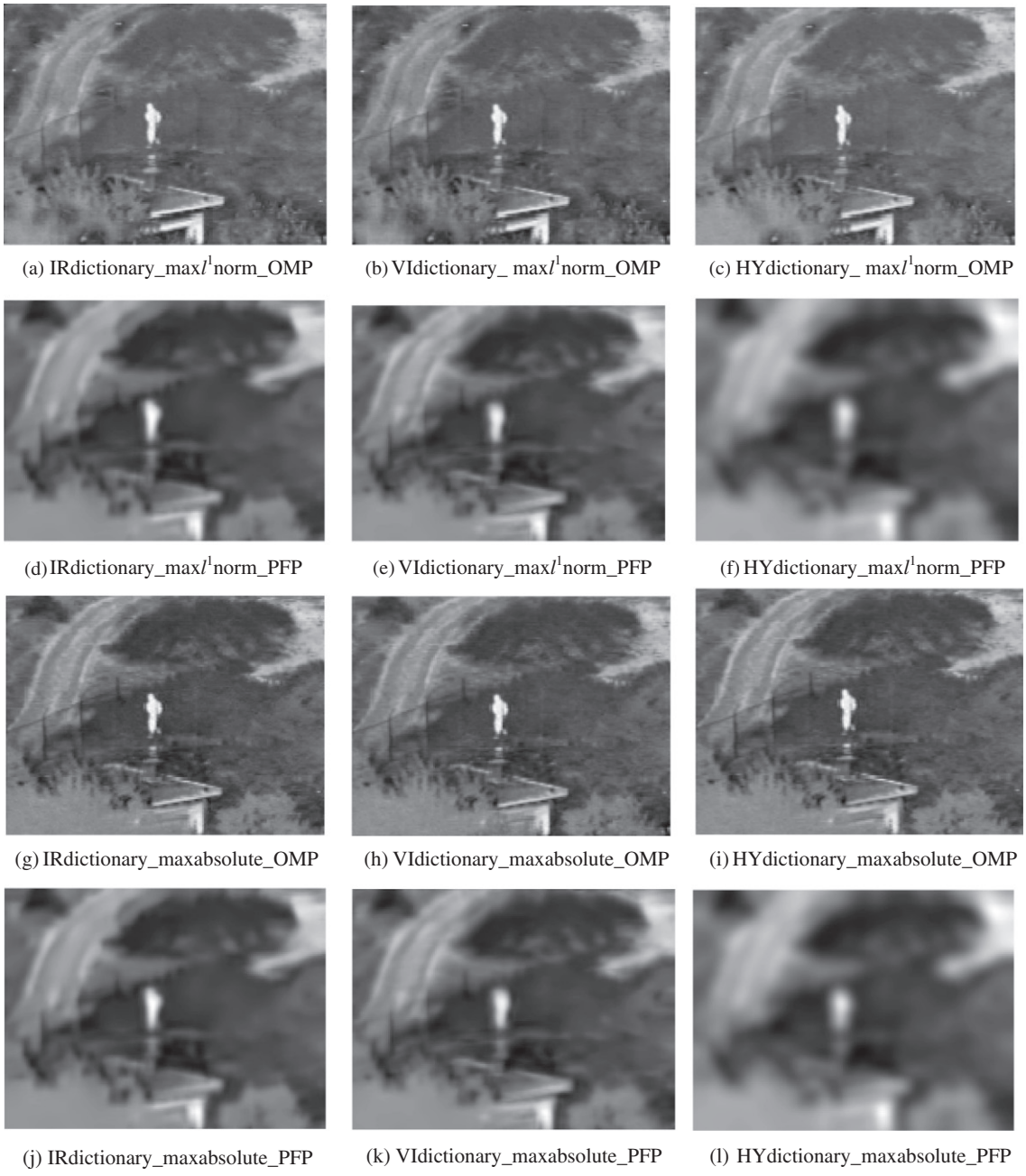


Fig. 8. Fused images of source images 3 using different approaches.

Table 1
Quantitative assessment of various fusion approaches for Fig. 6.

Fig.6	a	b	c	d	e	f	g	h	i	j	k	l
MI	1.6001	1.5751	1.6518	1.4387	**1.5551**	1.4984	**2.3624**	2.1504	2.3528	1.4387	1.5551	1.4984
Q_W	0.5724	**0.5738**	0.5621	0.2512	**0.3262**	0.2133	0.5616	0.5635	0.5473	0.2512	0.3262	0.2133
$Q^{AB/F}$	0.5754	0.5638	0.5689	0.2769	**0.3084**	0.2617	**0.6218**	0.6048	0.6270	0.2769	0.3084	0.2617

number of atoms of dictionaries shown in Fig. 4 is equal to 256. Fig. 10 demonstrates two dictionaries of different atom numbers ($K = 128$ and $K = 384$), and fused images using those two dictionaries and the method *IRdictionary_maxabsolute_OMP*. Since K is equal to the length of the sparse vector, the large K results in burdensome calculation. Compared with quantitative assessment of fused image using $K = 256$ dictionary, different values of K have no obvious effect on the objective evaluations. Therefore, we can select the value of K only according to the criterion of $K \geqslant n^2$, $n \times n$ is the size of image patch used for dictionary training. In addition, Fig. 11

Table 2
Quantitative assessment of various fusion approaches for Fig. 7.

Fig.7	a	b	c	d	e	f	g	h	i	j	k	l
MI	1.4189	1.3258	1.6679	**2.0942**	2.0869	2.0010	**3.2490**	2.9635	3.1216	2.0942	2.0869	2.0010
Q_W	0.3426	0.3447	0.3541	0.1928	**0.2303**	0.1758	**0.4650**	0.4586	0.4600	0.1928	0.2303	0.1758
$Q^{AB/F}$	0.4450	0.4347	0.4694	0.2566	**0.2869**	0.2423	**0.5896**	0.5735	0.5828	0.2566	0.2869	0.2423

Table 3
Quantitative assessment of various fusion approaches for Fig. 8.

Fig.8	a	b	c	d	e	f	g	h	i	j	k	l
MI	1.7893	2.4241	1.9725	**2.0942**	2.0869	2.0010	2.7291	1.7589	**2.8064**	2.0942	2.0869	2.0010
Q_W	0.4822	0.4936	0.4820	0.2702	**0.2964**	0.1699	**0.4908**	0.4830	0.4793	0.2702	0.2964	0.1699
$Q^{AB/F}$	0.4929	0.4939	0.4965	0.1952	**0.2044**	0.1296	0.5154	0.4771	**0.5255**	0.1952	0.2044	0.1296

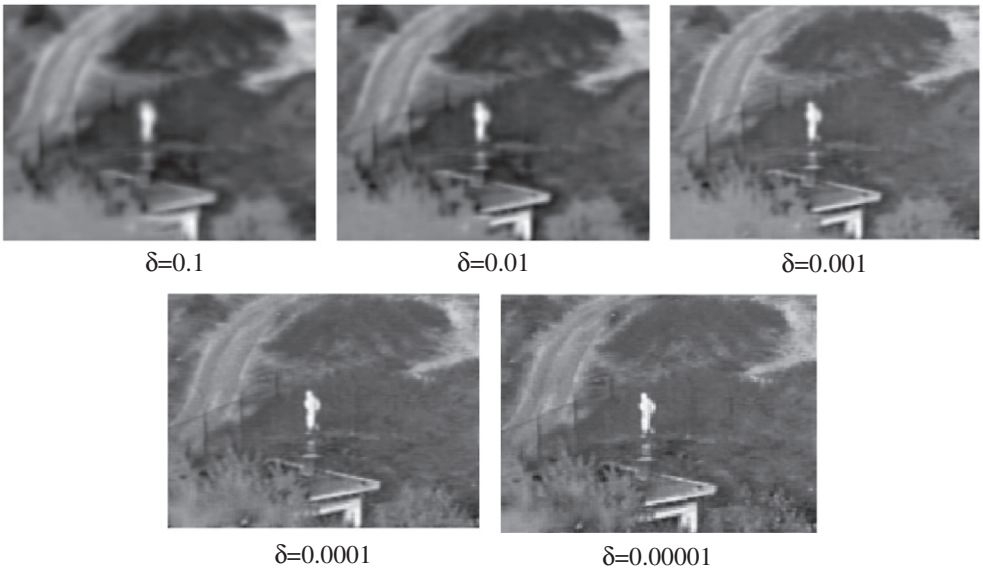


Fig. 9. Fused images according to different δ using BP algorithm.

Table 4
Quantitative assessment of various fusion approaches for Fig. 9.

δ	0.1	0.01	0.001	0.0001	0.00001
MI	2.1036	2.1711	**2.1941**	1.7647	1.6831
Q_W	0.3214	0.3731	0.4371	**0.5028**	0.4935
$Q^{AB/F}$	0.2299	0.3245	0.432	0.4518	**0.4760**

shows the overcomplete dictionaries of different iterations (200 and 300), fused images using these two dictionaries and the method *IRdictionary_maxabsolute_OMP* and correspondence objective evaluations. From Fig. 11 and compared with column g of Table 3, we can see that the objective evaluations have no obvious change when the iteration for dictionary training increases.

5. Conclusion

In this paper, we study 12 different approaches of image fusion based on CS theory according to different trained overcomplete dictionaries including the dictionary only using patches from the infrared images, the dictionary only using patches from the visible images and the dictionary using the combined patches, different sparse vector approximations containing OMP and PFP algorithms, and different fusion rules covering maximum l^1-norm and maximum absolute of entry of sparse vector which is firstly proposed by this paper. Three pairs of source images consisting of visible and infrared images are used to test the performance of the different methods. The objective evaluations containing **MI**, $\mathbf{Q_W}$, and $\boldsymbol{Q^{AB/F}}$ are used to quantitatively evaluate the fusion performances. The experimental results of these 12 methods can be concluded into the following aspects: (1) The method using OMP approximation can provide better fusion results in the condition of the same parameter setting. (2) The method *IRdictionary_maxabsolute_OMP* takes almost all the largest objective evaluations. (3) Using PFP approximation and the same dictionaries, the objective evaluations of fused images based on different fusion rules are identical. (4) As the value of δ decreases, the visual quality of the fused image using PFP approximation has been improved obviously. (5) Atom number

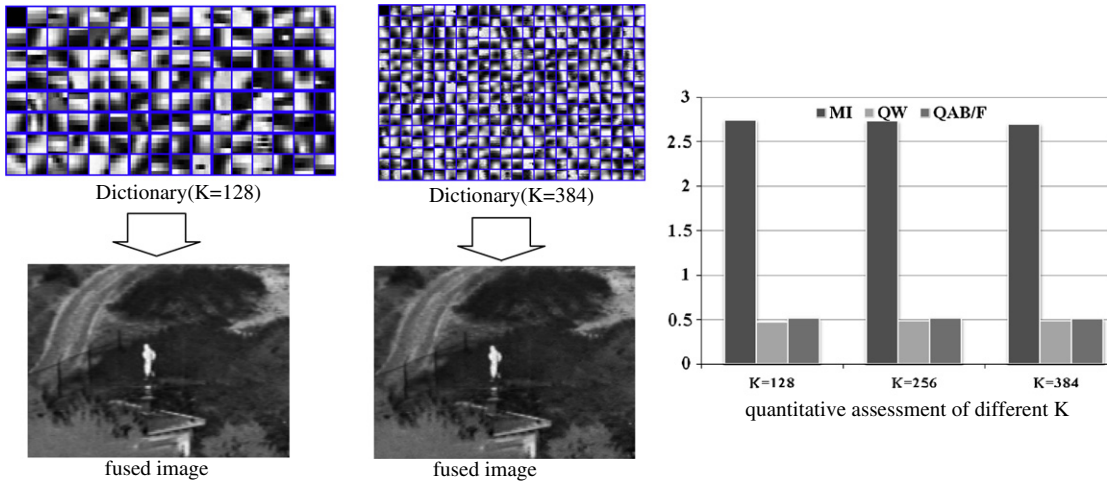


Fig. 10. Fused images and objective evaluations using dictionaries of different atom numbers.

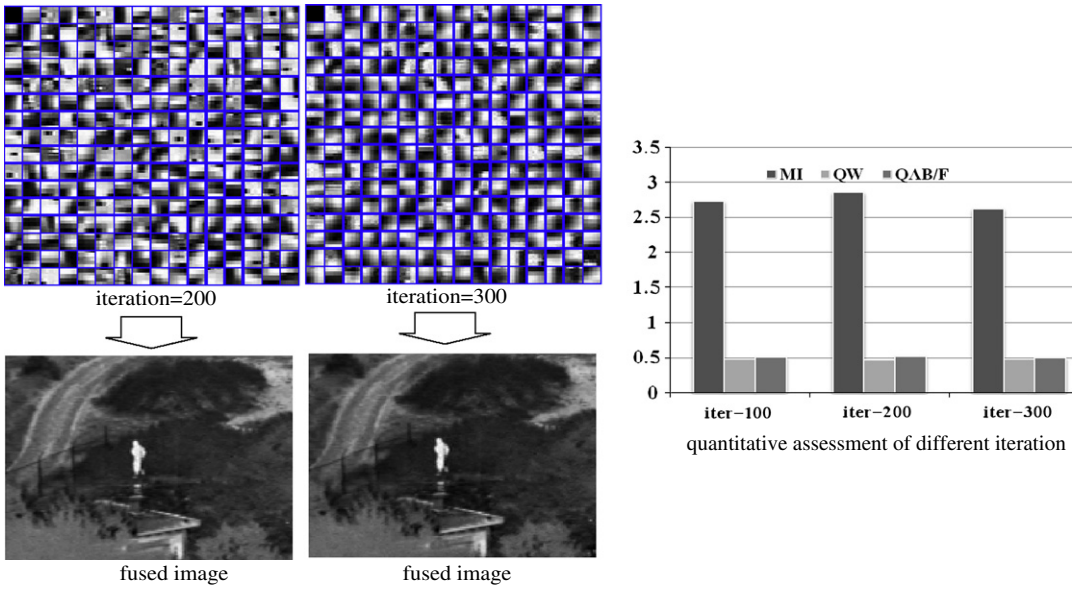


Fig. 11. Fused images and objective evaluations using two dictionaries of different iteration.

K contained in trained dictionary has no obvious effect on the objective evaluations. (6) The objective evaluations have no obvious change when the iteration for dictionary training increases.

Acknowledgement

The authors are grateful to www.imagefusion.org for providing data. They also like to thank the anonymous reviewers for their critical and constructive review of the manuscript. This study was co-supported by the National Natural Science Foundation of China (No. 61203170), the Fundamental Research Funds for the Central Universities ofChina (No. NS2012026), and Startup Foundation for Introduced Talents of Nanjing University of Aeronautics and Astronautics (No. 1007-YAH10047).

References

[1] N.N. Yu, T.S. Qiu, Fusion technology of infrared and visible images in compressive sensing, Signal Processing 28 (5) (2012) 692–698.
[2] X. Li, S.Y. Qin, Efficient fusion for infrared and visible images based on compressive sensing principle, IET Image Processing 5 (2) (2011) 141–147.
[3] A.A. Goshtasby, S.G. Nikolov, Image fusion: advances in the state of the art, Information Fusion 8 (2) (2007) 114–118.
[4] F. Mendoza, R.F. Lu, H.Y. Cen, Comparison and fusion of four nondestructive sensors for predicting apple fruit firmness and soluble solids content, Postharvest Biology and Technology 73 (2012) 89–98.
[5] S.T. Li, B. Yang, A new pan-sharpening method using a compressed sensing technique, IEEE Transactions on Geoscience and Remote Sensing 49 (2) (2011) 738–746.
[6] X.Y. Qian, Y.J. Wang, B.F. Wang, Fast color contrast enhancement method for color night vision, Infrared Physics and Technology 55 (1) (2012) 122–129.
[7] G. Piella, A general framework for multiresolution image fusion: from pixels to regions, Information Fusion 4 (4) (2003) 259–280.
[8] N. Mitianoudis, T. Stathaki, Pixel-based and region-based image fusion schemes using ICA bases, Information Fusion 8 (2) (2007) 131–142.
[9] B. Yang, S.T. Li, Multifocus image fusion and restoration with sparse representation, IEEE Transactions on Instrumentation and Measurement 59 (4) (2010) 884–892.
[10] H.T. Yin, S.T. Li, L.Y. Fang, Simultaneous image fusion and super-resolution using sparse representation, Information Fusion (in press) http://dx.doi.org/10.1016/j.inffus.2012.01.008.
[11] S.T. Li, J.T. Kwok, Y.N. Wang, Combination of images with diverse focuses using the spatial frequency, Information Fusion 2 (3) (2001) 169–176.
[12] G. Bhatnagar, Q.M.J. Wu, An image fusion framework based on human visual system in Framelet domain, International Journal of Wavelets Multiresolution and Information Processing 10 (1) (2012) 12500021–12500030.
[13] G. Pajares, J. Cruz, A wavelet-based image fusion tutorial, Pattern Recognition 37 (9) (2004) 1855–1872.

[14] W.C. Wang, F.L. Chang, A multi-focus image fusion method based on Laplacian pyramid, Journal of Computers 6 (12) (2011) 2559–2566.
[15] B. Yang, S.T. Li, Multi-focus image fusion using watershed transform and morphological wavelet clarity measure, International Journal of Innovative Computing, Information and Control 7 (5A) (2011) 2503–2514.
[16] V.S. Petrovic, C.S. Xydeas, Gradient-based multiresolution image fusion, IEEE Transactions on Image Processing 13 (2) (2004) 228–237.
[17] A.Z. Chitade, S.K. Katiyar, Multiresolution and multispectral data fusion using discrete wavelet transform with IRS images: Cartosat-1, IRS LISS III and LISS IV, Journal of the Indian Society of Remote Sensing 40 (1) (2012) 121–128.
[18] B. Yang, Z.L. Jing, Image fusion using a low-redundancy and nearly shift invariant discrete wavelet frame, Optics Engineering 46 (10) (2007). 107002-1-107002-10.
[19] S.T. Li, J.T. Kwok, Y.N. Wang, Discrete wavelet frame transform method to merge Landsat TM and SPOT panchromatic images, Information Fusion 3 (1) (2002) 17–23.
[20] G.Q. Chen; Y.H. Gao, Multisource image fusion based on double density dual-tree complex wavelet transform, in: 9th International Conference on Fuzzy Systems and Knowledge Discovery, IEEE, Sichuan, China, 29–31 May, 2012, pp. 1864–1868.
[21] J.J. Lewis, R.J. Ocallaghan, S.G. Nikolov, Pixel- and region-based image fusion with complex wavelets, Information Fusion 8 (2) (2007) 119–130.
[22] K. Liu, L. Guo, J.S. Chen, Image fusion algorithm based on finite ridgelet transform and cycle spinning, Journal of Jilin University (Engineering and Technology Edition) 40 (4) (2010) 1075–1080.
[23] Z.F. Shao, J. Liu, Q.M. Cheng, Fusion of infrared and visible images based on focus measure operators in the curvelet domain, Applied Optics 51 (12) (2012) 1910–1921.
[24] M. Elad, M. Aharon, Image denoising via sparse and redundant representations over learned dictionaries, IEEE Transaction on Image Processing 15 (2) (2006) 3736–3745.
[25] V. Cevher, A. Sankaranarayanan, M.F. Duarte, D. Reddy, R.G. Baraniuk, R. Chellappa, Compressive sensing for background subtraction computer vision, in: ECCV 2008, Lecture Notes in Computer Science, vol. 5303, 2008.
[26] J. Wright, A.Y. Yang, A. Ganesh, S.S. Sastry, Y. Ma, Robust face recognition via sparse representation, IEEE Transactions on Pattern Analysis and Machine Intelligence 31 (2) (2009) 210–227.
[27] J. Li, J.Z. Wang, W. Shen, Moving object detection in framework of compressive sampling, Journal of Systems Engineering and Electronics 21 (5) (2010) 740–745.
[28] H. Rauhut, K. Schnass, P. Vandergheynst, Compressed sensing and redundant dictionaries, IEEE Transactions on Information Theory 54 (5) (2008) 2210–2219.
[29] B. Yang, S.T. Li, Pixel-level image fusion with simultaneous orthogonal matching pursuit, Information Fusion 13 (1) (2012) 10–19.
[30] M. Aharon, M. Elad, A.M. Bruckstein, The K-SVD: an algorithm for designing of overcomplete dictionaries for sparse representations, IEEE Transactions on Image Processing 54 (11) (2006) 4311–4322.
[31] S.G. Mallat, Z.F. Zhang, Matching pursuit in a time frequency dictionary, IEEE Transactions on Signal Processing 4 (12) (1993) 3397–3415.
[32] S.S. Chen, D.L. Donoho, M.A. Saunders, Atomic decomposition by basis pursuit, SIAM Review 43 (1) (2001) 129–159.
[33] D.L. Donoho, M. Elad, V.N. Temlyakov, Stable recovery of sparse overcomplete representations in the presence of noise, IEEE Transactions on Information Theory 52 (1) (2006) 6–18.
[34] E. Candes, J. Romberg, T. Tao, Robust uncertainty principles: exact signal reconstruction from highly incomplete frequency information, IEEE Transactions on Information Theory 52 (2) (2006) 489–509.
[35] J.A. Tropp, A.C. Gilbert, Signal recovery from random measurements via orthogonal matching pursuit, IEEE Transactions on Information Theory 53 (12) (2007) 4655–4666.
[36] D. Needell, R. Vershynin, Greedy signal recovery and uncertainty principles, in: Proceedings of the Conference on Computational Imaging, vol. 6814, SPIE, San Jose, USA, 2008, pp. 68140J–68140J-12.
[37] D. Needell, J.A. Tropp, CoSaMP: iterative signal recovery from incomplete and inaccurate samples, Applied and Computational Harmonic Analysis 26 (3) (2008) 301–321.
[38] S.B. Chen, D.L. Donoho, M.A. Saunders, Atomic decomposition by basis pursuit, SIAM Journal on Scientific Computing 20 (1) (1998) 33–61.
[39] A.Y. Yang, S.S. Sastry, A. Ganesh, Y. Ma, A review of fast ℓ1-minimization algorithms and an application in robust face recognition, in: 17th IEEE International Conference on Image Processing (ICIP), IEEE, Berkeley, CA, USA, 2010, pp. 1849–1852.
[40] M.D. Plumbley, Recovery of sparse representations by polytope faces pursuit, in: Proceedings of International Conference on Independent Component Analysis and Blind Source Separation, Springer-Verlag, Charleston, SC, USA, Berlin, 2006, pp. 206–213.
[41] A. Gersho, R.M. Gray, Vector Quantization and Signal Compression, Kluwer Academic, Norwell, MA, 1991.
[42] Z. Liu, D.S. Forsyth, R. Laganiere, A feature-based metric for the quantitative evaluation of pixel-level image fusion, Computer Vision and Image Understanding 109 (1) (2008) 56–68.